DOCUMENT RESUME

ED 109 242 TM 004 705

AUTHOR Doherty, Victor W.
TITLE Appropriate and Inappropriate Uses of Performance Standards in School Evaluation.
PUB DATE [Apr 75]
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Criterion Referenced Tests; Educational Objectives; *Evaluation; *Evaluation Criteria; Norm Referenced Tests; *School Systems; *Standards; *Student Evaluation

ABSTRACT

The logic, i.e. the purposes, assumptions and implications of using performance objectives in evaluating schools is reviewed. The distinctions are recognized between standards for individuals and standards for groups as well as the distinction between normative and criterion-referenced standards. It is not appropriate to set standards for types of learning where cognitive development is a principal contributor to the type of capability being measured, if the intent of those standards is to try to get all students to achieve at a specified level. The integrated skills of reading and mathematics problem solving are examples. It is appropriate to set standards for acquiring finite bodies of information, or skills where their acquisition at the level designated is judged to be necessary for a well-defined purpose such as job competence, personal competence, social competence, or learning prerequisite to higher levels of education or training. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED109242

Appropriate and Inappropriate Uses of Performance Standards in School Evaluation

by

Victor W. Doherty
Portland, Oregon Public Schools

Use of performance objectives and "minimum standards" in school systems has become commonplace in the past few years, and it is a little surprising that the practice has not been subjected to closer examination by educational researchers. Support or condemnation of these practices rests on logical rather than empirical grounds. The analysis that follows is review of the logic, i.e., the purposes, assumptions and implications of various types of standards in use.

The first distinction recognized in this analysis is the difference between standards for individuals and standards for groups. The second distinction, which is superimposed on the first, is normative vs. criterion-referenced standards. There are, therefore, basically four cases:

Individual Standards:	Normative
Individual Standards:	Criterion referenced
Group Standards:	Normative
Group Standards:	Criterion referenced

The term "criterion referenced standard" is here defined as one in which a ratio or percent of items relating to a goal, sub-test, or test is designated as the acceptable level of performance.

The term "normative standard" is defined as one in which a point on a normative scale of a standardized test or sub-test is designated as the acceptable level of performance.

These definitions should be considered as independent of terminology applied to tests (often with confusion), such as "norm-referenced", "content referenced" and "criterion referenced".

The analysis that follows examines the four different types of standards referred to above. It includes illustrative "standards" statements and discusses the purposes, assumptions, and implications of each type.

TM 004 205

The chief difference between group and individual standards is the focus of the former on instructional effectiveness (setting a group standard will normally represent a desire to increase overall performance by increasing mean performance or the percentage of a group performing at the level specified) and the focus of the latter on a level of performance believed essential for every student to attain for some designated reason.

Relationship of Standards to Types of Learning

The setting of standards appears to involve not only a consideration of the types of standards it is possible to set, and the assumptions and implications of each, but also of the different kinds of learning provided for in the schools. The following analysis is provided to clarify the relationship between different kinds of learning and setting of standards of test performance. Different types of learning include:

Type 1. Acquiring a finite body of information with specific utility [e.g., acquiring information needed to hold a specific job, for personal safety, to do something (repair, construct, dismantle, cook, sew, etc.) to locate something.]

Type 2. Acquiring information of general utility (e.g., learning Avagardro's Law, the functions of societal institutions, the nature of matter, directions and locations).

Type 3. Learning a set of conventions, as in language, mathematics, music notation, accounting (e.g. knowledge of alphabet, correct word pronunciation, meaning of words, grammar, punctuation, signs and symbols of mathematics, algorithms, rules, definitions).

Type 4. Developing capabilities in the creative, effective use of conventions (e.g., developing ability to read with understanding, to write effectively, to apply mathematics in solving problems, to play or compose music).

Type 5. Developing values and concepts (e.g., learning to value something such as law, cooperation, truth; developing a concept such as gravity, supply and demand, capitalism).

Type 6. Learning manual skills (e.g. learning how to operate a lathe, a microscope, a band saw; learning how to sew, to operate a car, to dismantle a gun.)

The reason for providing examples of different kinds of learning is to help analyze the conditions under which the setting of standards is appropriate and inappropriate. If one examines the above types of learning, it will be noted that they tend to fall into two classes: developmental learning and mastery learning. Developmental learning characterizes Types 4 and 5, mastery learning, Types 1, 2, 3, and 6. Reading and math problem solving are developmental in that the interaction of cognitive processes with conventions of language and mathematics produces the total capability. While not discounting the necessity for students to acquire reasonable mastery of the conventions of reading and mathematics, the contribution of formal instruction in these conventions to the developmental capabilities which they undergird is probably far less important than the developing cognitive powers of the individual. If so, schools can take less and less credit for achievement gains in these developmental fields of learning as students advance through the grades. The setting of standards of mastery therefore appear to be appropriate when teachers want to determine if the conventions they are teaching are being learned; but less appropriate to apply to learning that is dependent on development and application of higher thought processes for which conventions are merely supporting tools.

In such cases, normative measures can be used to compare performance of individuals and groups against a standardizing group, but to set a "mastery" standard for all students is inconsistent with the knowledge that differential cognitive capacities will produce different levels of capability.

Similarly, conceptual or values learning is developmental. Given a particular value such as honesty or concept such as division of labor, it can be recognized that the information an individual acquires to create and support such values and concepts may vary a great deal, yet may functionally have very much the same meaning from person to person. Also, concepts and values are constantly conditioned and modified by additional information. Application of standards to such learning appear totally beyond any technology extant.

However, mastery standards may well be set for Type 1, 2, 3, and 6 learning. Here we are dealing with information or skills of finite, quantitatively measurable types. Here the question of appropriateness of applying standards does not focus on the nature of the learning, but on the need for a standard (is there a well defined need for mastery at a particular level?) or the impact on learners of setting a standard (will it be necessary to divert resources to slow learners that would normally be available to others to assure that all achieve the standard?) or the purpose of the standard (is it used to identify a group of students who should receive further instruction, as in reading or math?).

Summary

In summary, the setting of standards for individuals or groups should be preceded by an examination of the assumptions and implications in the foregoing analysis. It is not appropriate to set standards for types of learning where cognitive development is a principal contributor to the type of capability being measured, if the intent of those standards is to try to get all students to achieve at a specified level. The integrated skills of reading and mathematics problem solving are examples. It is appropriate to set standards for acquiring finite bodies of information, or skills where their acquisition at the level designated is judged to be necessary for a well-defined purpose such as job competence, personal competence, social competence, or learning prerequisite to higher levels of education or training.

Analysis of Types of Standards that may be Set for Individual or Group Performance on Achievement Tests

Individual Standards	Purpose	Assumptions and Implications
I. Normative (Each student will achieve a norm level regarded as minimally acceptable).		
a) "No student shall be more than (1 S.D.) below norm"	To insure a better performance than would normally occur from students who would fall below the designated point.	(A) An arbitrary designation of a point on a normative scale can be used to represent a defensible minimum achievement for every child.
b) "No students shall be more than (1 G.L.) below norm"		
c) "No student shall be under (25th) percentile on norm"		(I) To the degree imposed by the level of the standard, instructional time and resources would be required that would otherwise not be, or would be diverted from meeting needs of students who would normally exceed the standard.
d) "Each student shall score at grade level or above"		
e) "Each student shall score at mean or above"		
f) "Each student shall score at 50th percentile or above"		
II. Criterion Referenced		
a) Each student will correctly complete a specified percent of total items in the test - "Each student shall answer correctly not less than (100%, 90%, 75%, etc.) of all items in the test"	To help insure a level of mastery (of a total realm of knowledge and skills) considered necessary for a well-defined purpose.	(A) Defensible reasons exist for setting standards at an arbitrary point. Such reasons might be: (1) Not to fall below a level of performance that in the judgment of the instructor represents minimum performance required: a) as a prerequisite to more advanced learning in the same area; b) to perform functions on which the individual or society depends.
b) Each student will correctly complete a specified percent of items in each sub-test - "Each student shall answer correctly not less than (100%, 90%, 75%, etc.) of all items in each subtest"	To help insure a level of mastery (of each separate realm of knowledge and skills comprising the total realm of the test) considered necessary for a well-defined purpose.	

	Purpose	Assumptions and Implications
c) <u>Each student will correctly complete a specified percent of all items relating to each goal (objective) measured</u> - "Each student shall answer correctly not less than (100%, 90%, 75%, etc.) of all items relating to each goal (objective) measured"	To help insure a level of mastery (of each goal represented in the test) considered necessary for a well-defined purpose.	(I) Only under a continuous progress, individualized learning approach is it likely that such standards can be applied to individuals. Where students are pursuing basically the same curriculum at the same rate (and this is more often than not the case), setting such individual standards could only result in slowing the pace of instruction to a level detrimental to average and especially fast learners. The setting of any form of standa that could cause a shift in instr tional effort and priority should be carefully weighed for its possible damaging as well as constructive effect on the individua learner to which the standard is applied.

Groups Standards	Purpose	Assumptions and Implications
I. Normative		
a) The group will achieve a designated mean level of performance - "mean of group (class, grade) to be average (at grade level, 50th percentile, mean standard	To help assure that group mean performance attains the designated levels.	(A) That a defensible reason exists for the arbitrary standards imposed under each of the two applications shown at left.
b) A specified percentage of the group will achieve at or above a designated level of performance - "(100%, 90%, 75%, etc.) will score at or above average (at grade level, 50th percentile, mean standard score, 1 S.D. below norm, 25th percentile, etc.)"	To help assure that designated percentages of students atttain at or above a designated level on the test norms.	(I) Such a reason might be that the instructor has information regarding group performance on the same test, wants to improve his effectiveness, and sets a standard not lower and hopefully higher than that achieved by the prior group. Where a test has never been administered to a group before, the only basis for such a standard is an assumed capability on the part of the learners and the instructo Such standards can miss the mark widely and have little rational justification. If individual sta dards are necessary (see above) t satisfy occupational performance standards or levels of learning prerequisite to further learning, they would more logically be approached through the setting of individual rather than group standards.

II. Criterion Referenced	Purpose	Assumptions and Implications
a) Specified percentage of group will perform a specified percent of total test items correctly - "(100%, 90%, 75%, etc.) will complete correctly (100%, 90%, 75%, etc.) of all items in the test"	To help assure that designated percentages of students perform at designated levels of performance on the total test.	(A) Same as above (I) Same as above
b) Specified percentage of a group will correctly perform a specified percent of items on each subtest - "(100%, 90%, 75%, etc.) will correctly complete (100%, 90%, 75%, etc.) of all items in each subtest"	To help assure that designated percentages of students are performing at designated levels of performance on each student.	The principal virtue of the criterion referenced test applie to groups is to determine if specific goals of learning are being achieved by reasonable percentages of students, and if some goals are being more succes fully achieved than others. This requires analysis and interpretation of results more than the setting of standards.
c) Specified percentage of group will achieve a specified percent of correct items relating to each goal measured by the test - "(100%, 90%, 75%, etc.) of the students in the (class, grade) will correctly complete (100%, 75%, etc.) of the items relating to each goal measured by the test"	To help assure that designated percentages of students are performing at designated levels of performance on each goal (objective) represented by the items.	